English Language Use in Bilingual Primary Schools:

The Relation between Teacher Input and Children's Vocabulary Knowledge

Oana Costache (5653835)

Master's Thesis, Utrecht University

Linguistics: The Study of the Language Faculty

Supervisor: Prof. Dr. Rick de Graaff

Second reader: Dr. Elma Blom

Abstract

This study examined the relation between Dutch pupils' (age 6-7) vocabulary knowledge in English and their classroom-based English input provided by teachers. Using video-recorded observations of nine Bilingual Primary Education classrooms during English language instruction, teachers´ (N = 9) speech was transcribed and coded for lexical diversity, lexical sophistication, and syntactic complexity. Children's (N = 241) receptive vocabulary skills were assessed using the Peabody Picture Vocabulary Test. Results showed variation in children's vocabulary skills and teachers' language use. Bayesian hierarchical linear models revealed that after controlling for children's language scores in kindergarten and demographics, children's receptive vocabulary skills were positively related to teachers' use of diverse words and complex syntax, but not to teachers' total amount of talk. These findings suggest that exposure to high-quality classroom-based English promotes vocabulary learning of Dutch children enrolled in bilingual programs.

Keywords: early bilingual education, teacher language input, receptive vocabulary, English second language, Bayesian multilevel modeling

Table of contents

Introduction

Vocabulary knowledge has been identified as an important proxy for children's oral language skills, and it is also a particularly critical source of variation in the reading comprehension performance of first language (L1) and second-language (L2) learners (Lervåg and Aukrust 2010; Rydland, Aukrust, and Fulland 2012). In accounting for individual differences in vocabulary rates at the early stages of development, a growing body of research implicates the variation in the amount and diversity of children's early linguistic experiences at home (see review in Hoff 2006). In addition, the language children are exposed to in their preschool classrooms, by way of teachers' language input, has been shown to influence their early language development (Bowers and Vasilyeva 2011; Dickinson and Porche 2011; Gámez and Lesaux 2012; Gámez and Levine 2013; Gámez et al. 2017; Huttenlocher et al. 2002). Because of the positive effects of early language input on language development (Huttenlocher et al. 1991), recent recommendations for promoting children's second language learning include early rich exposure to an L2 environment (Celaya 2012).

In the Netherlands, these recommendations are currently being tested within a pilot study commissioned by the Ministry of Education and a consortium of researchers from the University of Maastricht, Radboud University Nijmegen, Expertisecentrum Nederlands and the Utrecht University that was implemented at the beginning of the school year 2014/2015. The pilot is set in the context of a particular bilingual education model called *tweetalig primair onderwijs* (TPO) currently being conducted in 19 Dutch primary schools for a duration of five years. Its main aim is to test the effects of immersion to the English language for 30-50% of the school day on pupils' language proficiency in Dutch and English (EP Nuffic 2016).

Several meta-analyses have been conducted on the effectiveness of bilingual education (e. g., (Francis, Lesaux, and August 2006; Greene 1997; Rolstad, Mahoney, and Glass 2005; Willig 1985)

and while the conclusions drawn from these reviews have been diverse (Rossell and Baker 1996; Willig 1987), the empirical evidence suggests that sustained bilingual exposure promotes later L2 development (Goldenberg 2008) and helps maintain L1 language skills (Genesee 2004). However, these studies focused on global measures of language use (e.g., comparisons between program types) and do not specify the features of the classroom language environment that help account for beneficial L2 effects. Developing a better understanding of the optimal language environment for second language learners requires a description of the type of language (i.e., quality) to which they are exposed, in addition to the amount of exposure to the language (i.e., quantity). The role of input may, indeed, be analogous in L2 and monolingual children in that the sheer amount of talk could be an important predictor for young emergent L2 children (Bowers and Vasilyeva 2011). On the other hand, it is unclear how exposure to qualitatively different linguistic input impacts on the vocabulary knowledge of L2 learners.

It is therefore important to identify how the talk children are exposed to early on in schools, as well as the demography they are embedded in, may support their L2 vocabulary development. Thus, the objective of the present study is to explore the quantity and quality of classroom-based English exposure of children attending bilingual primary schools in the Netherlands and examine the impact of this exposure to their receptive language outcomes. The findings may result in practical implications on language teaching designs in TPO schools, as well as for the training programs of TPO teachers.

Theoretical Framework

Before discussing the theoretical background to this study, a brief explanation of the use of some key concepts is given. Firstly, a distinction is often made between ‘foreign’ and ‘second’ language learning. Briefly, a ‘second’ language has social functions within the community where it is learned, whereas a ‘foreign’ language is learned primarily for contact outside one’s own community

(Littlewood 1984). In the literature, both terms are sometimes used interchangeably, however, in this study, the term *second language* will be used as a cover term for both 'foreign' and 'second' language, given the specific situation in the Netherlands where the English language is very prominent in society. Another distinction which is sometimes made is that between 'learning' and 'acquisition'. Learning refers to conscious processes for internalizing a second language, whereas acquisition refers to subconscious processes (Littlewood 1984). Littlewood uses the term acquisition when referring to a process when children acquire their first language, otherwise he uses the term learning. Following him, in this study, learning will be used as a general term, except when the distinction is crucial to immediate discussion.

Early second language learning

One of the most controversial issues in the field of second language learning is the question of age. The impact of age on ultimate attainment is an issue of continuing dispute (Scovel 2000), yet it is widely accepted that "younger is better", in that children tend to learn languages with more ease, and with a more successful ultimate attainment than do adults (Birdsong 1999; Pinker 1994). However, the exact duration of this sensitive period is debatable. DeKeyser (2000) points to results indicating that 'somewhere between the ages of 6-7 and 16-17, everybody loses the mental equipment required for the implicit induction of the abstract patterns underlying a human language' (DeKeyser 2000, 518). As a consequence of this cognitive change, younger children are most effective in second language learning when being exposed to implicit linguistic input around them, as opposed to explicitly being taught grammar rules and vocabulary.

Bilingual education paradigms

With growing evidence showing that learning a second language at an earlier age could prove more efficient for the ultimate attainment, many educational systems throughout the world have

started to implement the instruction of several languages from an early age in their curriculum (Genesee 1985; Lindholm-Leary and Genesee 2014). However, some educationalists have quickly realized that teaching languages the traditional way, that is, in a rather formal context at specific times during the week did not provide children with the communicative proficiency they would need to function and succeed in the L2-speaking community (Genesee 1985; Lambert and Tucker 1972).

Immersion programs

In order to remedy this, educators and psychologists in Canada set up *immersion programs* in the second half of the last century. In this approach, the second language is not simply taught as another subject in the curriculum, but rather is the medium through which the curriculum itself is taught (see Genesee [1983] for a complete description of the Canadian immersion programs). Immersion programs can be distinguished according to whether they offer *early* immersion in the second language, beginning in kindergarten, *delayed* immersion, beginning in grade 4 or 5, or *late* immersion, beginning in grade 7 or 8. A distinction is also made between *total* immersion alternatives, which offer all or nearly all instruction during one or more school years in the second language, and *partial* immersion alternatives, which usually provide 50% of instruction in English and 50% in the second language. Evaluation results indicate that students in immersion programs attain high levels of functional proficiency in the second language (e.g., reading and listening comprehension, oral communication, and writing), although they seldom attain truly native-like mastery of the elements and rules of the language (Genesee and Lindholm-Leary 2014; Lyster and Genesee 2013).

Content and language integrated learning

In Europe, Canadian immersion has served as an important role model for a bilingual educational teaching approach that was launched in the 1990s due to the increasing need for higher level of foreign language proficiency by pupils (Dalton-Puffer, Nikula, and Smit 2010). However, the

effort to implement the Canadian immersion model into the European bilingual education was not particularly successful. Marsh (2002, 56) comments that the researchers found out that 'immersion bilingual education was successful for majority language speakers (e.g. in Quebec) more than for those coming from a minority language background'. The educational driver was to design and adapt language teaching approaches like those in Canada so as to provide a wide range of L1 students with higher levels of L2 competence.

During the 1990s, the acronym CLIL became the most widely used term for the *content and language integrated learning* in Europe. Coyle, Hood, and Marsh (2010) define CLIL as 'a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language' (2010, 1). In other words, CLIL involves an explicit focus on language as an object of teaching during content lessons (Coyle, Hood, and Marsh 2010). A key difference between CLIL and immersion methodologies is the explicit emphasis on the integration and equilibrium of language and content (Cenoz 2013). According to Llinares, Morton, and Whittaker (2012), the majority of CLIL programs in Europe focus on English and most of them are implemented in secondary schools, although there are also examples of CLIL at primary and tertiary levels (Hüttner and Rieder-Bünemann 2007; Núñez and Dafouz 2007).

The Dutch Context: Tweetalig Primair Onderwijs (TPO)

In the Netherlands, CLIL has become a popular teaching practice and is incorporated into the so-called 'tweetalig onderwijs' (TTO, bilingual education). TTO originates from the late nineties when secondary schools decided to offer a more international form of education as a reaction to Dutch parents' demands. Recently, several primary schools joined a project led by EP-Nuffic, with the aim of introducing a form of bilingual primary education (*tweetalig primair onderwijs*, TPO) in the Netherlands. The aim is to allow schools to respond to a growing demand from parents for earlier and

more foreign language instruction. The program assumes partial immersion in the English language for between 30-50 percent of the school day (EP Nuffic 2016), similarly to the early immersion methodology in Canada.

Individual differences in second language learning

Factors influencing L2 learning rates that vary among individuals can be categorized as internal or external to the learner (Ellis 2004; Paradis 2011). Child-internal factors include the following: gender, socioeconomic status (SES), language aptitudes, and cognitive maturity as represented by chronological age. Prior research indicates that each of these factors can potentially influence children's L2 acquisition rates (c.f. Cook 2016; Paradis 2011). While teachers can hardly influence the impact of child-internal factors, they do have some power over factors that come from outside the individual.

Child external factors are mainly factors that determine the quantity and the quality of the input the child receives in the target language (Paradis 2011). A child's quantity of input in the L2 can vary based on overall length of exposure at home, or differences in exposure time in school. Regarding quality of the input, variation in experience with (non)native-speaker input, or rich and complex input could all contribute to differences in children's L2 acquisition rates (Unsworth et al. 2014). Previous studies that have examined the impact of child-external factors at home (De Houwer 2007; Paradis et al. 2011; Quiroz, Snow, and Jing Zhao 2010; Scheele, Leseman, and Mayo 2009) and in school (Vermeer 2001) have touched on both quantity- and quality-oriented factors, and have found both to have an impact on children's development. However, few existing studies have examined both external input factors alongside a set of internal factors to assess the relative contribution of each type of individual difference factor (for an overview see Unsworth 2016).

The role of linguistic input on vocabulary development

A growing body of literature shows that children exploit their linguistic input in the service of word learning (e.g. Huttenlocher 1998; Huttenlocher et al. 1991; Weisleder and Fernald 2013; Weizman and Snow 2001). However, whereas studies with monolingual children have focused on both the quantity and quality of language input (see [Hoff 2006a] for a review), the literature on L2 learners has focused primarily on the quantity of language input (see Unsworth 2016).

Research with monolinguals

Research on talk exposure and monolingual children's oral language outcomes, particularly their vocabulary, suggests that the sheer amount of talk children hear may promote acquisition. Purely quantitative factors, such as the frequency with which a child hears a new word, is a major factor explaining differences in vocabulary size for young monolingual children (Hoff and Naigles 2002). Several studies have also examined the impact of exposure to rare or sophisticated words in mother-child conversations, finding that children's vocabulary growth related to the occurrence of sophisticated lexical items in input (Weizman and Snow 2001). For example, Weizman and Snow found that sophisticated input (exposure to words that were not within the 3000 most common) was a better predictor of vocabulary performance than quantity of lexical input overall.

The few studies focused on linguistic input in classrooms show that the variation in the quality of teacher input accounts for differences in school-aged monolinguals' language skills, such as vocabulary and grammar (Aukrust 2007; Dickinson and Porche 2011; Huttenlocher et al. 2002). For instance, Huttenlocher et al. (2002) found that preschoolers (age 5) exposed to teachers who used more syntactically complex utterances demonstrated greater syntactic ability compared to preschoolers exposed to teachers who used less syntactically complex utterances. In addition, Dickinson and Porche (2011) found that monolingual children's (age 6) exposure to sophisticated vocabulary in preschool

predicted fourth-grade reading comprehension, with effects mediated by children's kindergarten vocabulary and literacy.

Studies with middle school children further suggest that children's gains in oral language skills are related to the lexical complexity and the number of unique word types in teachers' input (Gámez and Levine 2013). Interestingly, Gámez and Lesaux (2012) found no significant relation between middle school students' (age 11.5) English vocabulary skills and their teachers' total amount of English input. However, they found a link between students' English vocabulary skills and their teachers' grammatically complex and lexically diverse speech.

Research with second language learners

Very few studies addressed how school talk exposure may predict second language vocabulary. Among the exceptions is a study on vocabulary growth in preschool children (age 5) learning English as L2 in a monolingual English-speaking classroom by Bowers and Vasilyeva (2011). The authors reported that the overall amount of speech produced by teachers was a significant predictor of growth in children's vocabulary scores within one school year. The lexical diversity (ratio of different words to the total number of words) of teacher talk was not related to L2 children's but to monolinguals' scores, whereas the Mean Length of Utterance (MLU) showed a negative relation whereby an increase in teachers' number of words per utterance corresponded to a decrease in L2 children's vocabulary growth. These findings suggest that exposure to a lot of talk is more supportive for second language learners' vocabulary growth, while the diversity of word types is a stronger predictor of the monolingual children's vocabulary growth.

Further, in a study of teacher-student conversations in multi-ethnic preschools serving high numbers of L2 (Norwegian) speakers, Aukrust and Rydland (2011) found that the overall vocabulary diversity of the preschool conversations predicted children's (age 5) L2 vocabulary skills in the first

grade, as did the density of words. However, the word density indicator included as part of this study measured the rate of words used by teachers and pupils during conversation. Thus, the finding that it predicted L2 learners' first grade vocabulary skills can also be interpreted as showing that a greater amount of interactive conversations, not just the total amount of teachers' L2 exposure, is a critical factor in promoting language development.

In contrast to Bowers and Vasilyeva (2011), Gámez (2015) and Gámez et al. (2017) found a positive relation between children's (age 6) English expressive skills and their teachers' syntactic complexity. These studies focused on Dual Language Learners (DLLs) enrolled in a Spanish-English bilingual education program (i.e., ELLs) and used the MLU in words (Gámez 2015) and multi-clause utterances (Gámez et al. 2017) as an indicator of syntactic complexity in English. Their results showed that children in classrooms where teachers used more syntactically complex English exhibited greater gains in English expressive language skills than did their peers in classrooms with teachers who used syntactically less complex speech. These studies thus underscore the importance of investigating how talk exposure in early education may scaffold the vocabulary learning of second-language learners in particular.

Bilingual education in the Netherlands

Historically, the question of whether bilingual instruction increases or reduces student learning opportunities has been controversial. Three studies conducted in the Netherlands (Admiraal, Westhoff, and de Bot 2006; van der Leij, Bekebrede, and Kotterink 2010; Verspoor, de Bot, and Xu 2015) examined the effectiveness of different bilingual program models with students of different ages (between 8 and 15 years). In all three studies, students in bilingual programs in English and Dutch significantly outperformed their peers on English and Dutch measures in schools where Dutch was the language of instruction and English was taught for four hours a week or less.

Admiraal et al. (2006) showed that students attending bilingual secondary education had higher scores for their English language proficiency in terms of oral proficiency and reading comprehension, however, no effects were found for receptive word knowledge. Importantly, Verspoor et al. (2015) demonstrated that L2 development is a dynamic process in which individual differences, especially the initial level of proficiency and scholastic aptitude, interact with other variables such as the amount of exposure to the language (both in and out of school).

Teacher language proficiency in TPO

Because early bilingual education in the Netherlands is a bottom-up development, with limited governmental regulation, there is much variety in the amount of classroom exposure and teachers' English language proficiency. In the majority of bilingual primary schools, classes are taught by a non-native- or near-native speaker class teacher, whose mastery of English ranges from B2 to C2 within the CEFR framework[1] (Driessen et al. 2016). In the Netherlands, an intermediate English proficiency (B1) at the end of teacher training colleges is desired, and a B2 level is preferred, while the guidelines for TPO indicate a C1 level as preferrable.

How this affects pupils' language development has not yet been subject to systematic investigation. One exception is a study by Unsworth et al. (2014) investigating the development of 4-year-old children's receptive vocabulary and grammar skills in English as a function of their participation in early English foreign language programs in the Netherlands. The researchers reported that teachers' language proficiency was a significant predictor of children's receptive vocabulary scores. Specifically, the children with a non-native teacher at CEFR-B level only scored significantly lower than children with a non-native teacher at the same level who co-taught with a native teacher, children with a native teacher only, and children with a near-native teacher (C1-C2).

[1]The Common European Framework of Reference for Languages (henceforth CEFR) helps to describe the level of proficiency on the basis of six broad levels. The levels range from basic (A1 and A2) through intermediate (B1 and B2) to advanced foreign language proficiency (C1 and C2) (Council of Europe 2001).

However, teacher proficiency involves much more than language proficiency alone. For example, factors such as teaching experience can have an influence on teachers' language use. For example, teachers have been shown to intentionally adjust the grammatical complexity of their speech and use higher-frequency vocabulary items in order to make classroom input comprehensible to students (for comprehensive overviews see Chaudron [1988] and Nikolov and Djigunovic [2011]). Furthermore, in a study of input provided by teachers in German bilingual preschool programs, Weitz, Pahl, Flyman Mattsson, Buyl, and Kalbe (2010) observed that quality of language input, as measured in terms of didactic skills and strategies employed, had a greater impact on children's linguistic development than the amount of input.

Because young L2 learners differ from their monolingual peers in terms of English skills, it is likely that there may be differences in the pattern of relationship between a particular aspect of input and their learning outcomes. For example, the overall amount of input, containing multiple repetitions of common words, may be important for L2 learners still acquiring basic English vocabulary. On the other hand, Carlo et al. (2004) indicate that second language learners in particular rely on the discursive context in which words appear. When words appear within diverse and complex discourse contexts, children are offered more opportunities to develop a rich understanding of the various meaning aspects of the word. For instance, a first encounter with a word might provide information about syntactic word class and some general specification of meaning domain, whereas subsequent encounters will expand the semantic specification and may lead to discovery of polysemous possibilities (Bloom 2002). A more diverse and structurally complex input could provide L2 learners with useful distributional regularities (Onnis 2012) or, on the contrary, it could hinder L2 learning due to cognitive overload (Kanokpermpoon 2013). These possibilities were closely investigated in this study.

Present study

The present study examined the role of teacher language input during English instruction in nine Dutch bilingual primary schools as a potential predictor of children's receptive vocabulary scores. Specifically, the goal was to test whether and how teachers' speech, that is, their amount of talk and use of a diverse set of sophisticated vocabulary and complex syntax, is related to children's vocabulary outcomes. To do so, teachers' amount of talk (i.e., quantity), use of sophisticated and diverse vocabulary, and syntactic complexity (i.e., quality) was analyzed from transcriptions of videotaped observations. Additionally, the performance of children on the Peabody Picture Vocabulary Test was documented as a measure of receptive vocabulary skill. To investigate the impact of teacher input on children's vocabulary performance, Bayesian hierarchical linear modeling was used, to account for the small sample size and for the nesting of children within classrooms.

Based on existing literature, it was hypothesized that amount of talk would be positively related to children's vocabulary scores (Bowers and Vasilyeva 2011). With regard to the quality of input, lexical diversity and sophistication was anticipated to have a positive relation with children's vocabulary knowledge (Aukrust and Rydland 2011; Bowers and Vasilyeva 2011; Gamez and Lesaux 2012; Gamez and Levine 2013), while no expectation was formulated for the syntactic complexity due to the mixed results in the literature (Bowers and Vasilyeva 2011; Gámez 2015; Gámez et al. 2017). In addition, parental education and home language have been shown in many studies to correlate highly with children's second language acquisition skills (Droop and Verhoeven 2003; Leseman 2000). These factors can introduce confounds into the analysis of relationships between classroom talk exposure and subsequent language knowledge and were therefore considered in the analysis, as were teachers' language proficiency and teaching experience in primary education which could also have an influence on teachers' language use (Unsworth et al. 2014). The following research questions were addressed:

(1) To what extent do children's characteristics such as *home language*, *socioeconomic status (SES), age* and *gender* relate to children's vocabulary knowledge?

(2a) Does the *amount of teacher speech* relate to children's vocabulary knowledge? (Quantity)

(2b) Does teachers' use of *lexically diverse words* relate to children's vocabulary knowledge? (Quality)

(2c) Does teachers' use of *low frequency words* relate to children's vocabulary knowledge? (Quality)

(3) Does teachers' use of *complex sentences* relate to children's vocabulary knowledge? (Quality)

Method

The data for the present study were collected as part of the research project Flankerend onderzoek Tweetalig primair onderwijs (FoTo), which has the goal to evaluate the progress of the pilot study (Driessen et al. 2016). The data were drawn from nine first grade[2] classrooms in nine schools across the Netherlands, which participated in the second wave of the FoTo project. The data were collected systematically (videotaped) in naturalistic classroom situations during English instruction. The current study targeted only the first part of the English lessons (each lasting approximately 20 minutes). The teachers were instructed to teach two lesson parts, one predesigned within the project (see Section on Lesson Design) and one in accordance with their normal teaching methods. Children's vocabulary receptive skills were assessed using the Peabody Picture Vocabulary Test (Dunn and Dunn 2007).

[2] In the Netherlands, preschool, Kindergarten and primary school are integrated in the Basisschool (Primary School) with grades (groepen: 'groups') 1–8 for 4–12 year olds. In this paper, the international equivalent is used with grade one as the first grade with formal instruction and practice of reading and other academic skills (group 3 of the Basisschool).

Participants

The teachers were recruited from schools as part of the regional collaboration project FoTo. All teachers participated voluntarily. The teachers and parents of the participating children gave informed consent before the start of the study, and all procedures conformed to existing ethical guidelines. To obtain measures on children demographics and home language exposure, parents were asked to fill out a questionnaire in either Dutch or English. Although the longitudinal pilot is currently being conducted in 19 schools, only the 12 schools that were evaluated at the start of the study were included in the second wave. Data from three of the 12 schools had to be excluded from analysis due to technical problems or teachers' unwillingness to participate.

Teachers

Nine primary education teachers corresponding to the nine participating first-grade classrooms were included in this study. The teachers were between 26 and 63 years old (M = 42.89, SD = 14.84) and they varied in their level of teaching experience, ranging from three to 40 years teaching in general primary education (M = 17.67, SD = 15.89) and one to 35 years teaching in English primary education (M = 8.44, SD = 10.58). All teachers reported holding a bachelor's degree or higher and having a state teaching credential. In addition, all teachers except for one had followed additional theoretical courses for teaching English and five teachers indicated that they had further working experience in the field (e.g. teaching internships abroad).

The language proficiency of teachers was determined by teachers' self-assessments and by an evaluation of the last part of the interviews conducted with them after the lessons, in which they had to answer a couple of questions in English. The self-assessments were compared against the language evaluations performed by a research assistant associated with the FoTo project. Based on this cross-checking, two teachers were native speakers of English, four were near native speakers (level C1 on

the scale of the Common European Framework of Reference for Languages, CEFR) and three were non-native speakers (level B1-B2 on CEFR).

Students

This study used data from 241 children of the 468 who composed the second cohort of the pilot evaluation and for whom outcome data were available. Forty-seven percent of these children were female, and the full group ranged in age from 6 to 8 years ($M = 6.75$, $SD = .35$). The number of children in each class ranged from 9 to 45 ($M = 31.06$). The socioeconomic status of the pupils was calculated as the average of both parents' education. Only 2.5% of the parents reported completing only a low vocational education or less, 7% reported at least some secondary or high school education (MAVO, VMBO), and 52% reported completing some higher education (university or applied sciences). Thirty-three and thirty-seven percent of the mothers and father, respectively, did not provide information about their education levels. The mothers of 46% of the children reported speaking Dutch as their native language, 3% were native speakers of English, 1% of Frisian and 16% of the mothers reported a different native language. Thirty-four percent did not provide information on their native language.

Materials and procedure

Children assessment

The Peabody Picture Vocabulary Test. Children's vocabulary was measured with the Peabody Picture Vocabulary Test— Fourth Edition (PPVT–IV). The children were tested by examiners working within the FoTo project. The first assessment (PPVT 1) took place between February and June 2015 and the second assessment (PPVT 2) was administered between February and June 2017. The PPVT-4 consists of 228 items, divided into 19 sets of 12 items based on increasing age-normed difficulty. Per item, there are four possible answer options, shown as pictures on a computer screen.

Children are given the target stimuli orally and asked to indicate the picture that represents the item heard. There are four possible answer options: A correct item and three distractors that are based on semantic resemblance to the correct word. The items broadly sample words that represent 20 content areas (e.g., actions, vegetables, tools) and parts of speech (nouns, verbs, or attributes) across all levels of difficulty (Dunn and Dunn 2007).

The testing procedure started with the set that matched the age of each child at the date of testing (basal item). The testing was stopped when the child reached the ceiling criteria, that is, when it indicated eight incorrect items out of a set of 12. Each testing session lasted about 20 minutes. Children's raw scores were calculated by subtracting the total number of errors made between the basal and ceiling item sets. These scores were used as the vocabulary outcome measure in the analyses.

Classroom observations

Lesson design. The teachers were instructed to give a regular and a pre-designed lesson on a subject of their own choice that had to incorporate an inside-outside circle activity. Regarding the regular lessons, the teachers were left entirely at liberty to decide on the content. The pre-designed lesson on the other hand consisted of three components: a classical instruction, a cooperative work activity (inside-outside circle activity), and a plenary round up. The aim of this lesson was to stimulate language production of children. In the inside-outside circle activity, half of the children in the class formed a circle, facing outward. The other children each stood facing one person in the circle. The pairs briefly either interviewed each other or reviewed previously introduced vocabulary. At the teacher's signal, the children stopped talking. The teacher asked one pair to repeat their conversation to the whole class, or ask what the other person said. Then the teacher gave instructions either for the inside or outside circle to rotate a few steps before forming new pairs.

The topics that the teachers chose for this activity were very diverse, nevertheless the duration was comparable across lessons. Table 1 presents an overview of the topics and the duration of the pre-designed lessons per teacher.

Table 1. Overview of lesson topics and duration per teacher

Teacher	Topic	Duration
A	Nature and the world	20:57
B	The clock	22:03
C	House and baby animals	22:02
D	Animals	21:20
E	Professions	22:00
F	Weather forecast	20:07
G	Fruits	21:12
H	Plants	22:55
I	House, spring, school	20:53

Transcription procedure

To control as much as possible for random variation in input that may occur in different instructional contexts, only the pre-designed lesson was transcribed following the Codes for Human Analysis of Transcripts conventions (MacWhinney 2000) because during this time the instructional content was comparable across lessons (see section on Lesson Design). The transcription procedure was stopped when the inside-outside circle activity ended, thus the duration ranged from 20 to 23 minutes ($M = 21.34$, $SD = 0.85$). The unit of transcription was the utterance. Utterance boundaries were determined based on turn taking, pauses, and intonation patterns (Brown 1973; Rowe 2012). Partly or completely unintelligible utterances were coded as such. No distinction was made between individual children, such that the speaker was identified as either a teacher or a child.

Entirely Dutch utterances were marked with the precode [- nld] (Macwhinney 2000). A mixed utterance was defined as an utterance containing elements from both languages (Cantone 2007). Mixing in the English context typically meant that the teacher was inserting Dutch elements into an English frame, as in *In Dutch we say Frans*. In these cases, the Dutch words were marked with a postcode '@s' indicating use of L2 because English was assumed to be the dominant language of the lesson. Furthermore, instances that could not be categorized either as English or as Dutch, for instance when the teacher used words such as *mandarine,* were marked with the postcode '@n' indicating a neologism.

Data analysis

Linguistic analyses

All transcripts were prepared for analysis using the Computerized Language Analysis software (CLAN), which is part of the Child Language Data Exchange System (CHILDES; MacWhinney 1995). An overview of the linguistic analyses and the methods used to obtain the language measures is provided in Table 2. The unit of analysis was teachers' English utterances. Filler sounds (e.g., um, hmm, sh) as well as exclamations and interjections (oh, hmm) were excluded from any analyses as was speech marked as unintelligible. Dutch utterances and words marked with a pre- or post-code and neologisms were removed from the analysis (but they were included in the calculation of the amount of talk).

Table 2. Overview of language measures with corresponding methods to obtain these measures

Measure	Method
Amount of talk	
Tokens	Command FREQ in CLAN
Utterances	Total number of utterances in English and Dutch
Lexical diversity	Guiraud's index
Lexical sophistication	Filtering out of common words
	Command FREQ in CLAN for updated list
	Number of teachers' sophisticated tokens relative to all tokens
Syntactic complexity	Number of teachers' complex utterances relative to all teacher utterances

Amount of talk

Teachers' amount of talk was calculated at the word and utterance level as the total number of words ($Tokens_{Total}$) and number of utterances ($Utterances_{Total}$) produced, respectively. A distinction was made between English and Dutch utterances and words, and the number of Dutch and English words and utterances was calculated for each teacher separately.

Lexical diversity

Traditionally, in child language research, lexical diversity has been measured using a type–token ratio (TTR). The term type refers to the number of unique words used in a text, in contrast to the term token, which refers to every instance a particular word occurs. However, because TTR seems to be dependent on the total number of tokens produced (Rietveld and Van Hout 1993; Van Hout and Vermeer 1988), other means of measuring lexical diversity were designed, including Guiraud's index (Guiraud 1960) and the D index (Malvern and Richards 2002). Although objections have been raised against all the available measures (see Torruella and Capsada 2013), in this study the Guiraud's index was used to measure lexical diversity because all transcripts met the requirement of a minimal sample

size of 250 tokens (Richards 1987). Guiraud's index is calculated by dividing the number of types by the square root of the number of tokens ($V/\sqrt{N}$). By taking the square root of the number of tokens, the problem of a negative correlation with increasing sample size (as with TTR) is avoided (Guiraud 1960).

Lexical sophistication

Sophisticated words – low-frequency words – were identified by filtering out high-frequency words from teachers' transcripts. The Target Word list (*streefwoordenlijst*) provided by EP-Nuffic of approximately 2,000 words that pupils attending BPE schools should know by the end of kindergarten was the starting point (EP-Nuffic, n.d.). The list was expanded by adding the derivationally inflected forms of the base words (e.g. *'s, s, es, ies, d, ed, ied, ing*) to create a list of 3,247 common word forms to which proper nouns found in the transcripts were added.

This list was then compared to the Dale-Chall easy word list consisting of 3,000 words understood by 80 % of English monolingual fourth-grade children (age 9–10) which has been previously used in studies investigating relations between language input and children vocabulary outcomes (Chall and Dale 1995). To calculate list similarity, a function available in Excel was used. The similarity of the two lists was 0.78. Although previous studies have used the Dale-Chall list as guidance for establishing lexical sophistication, this list is designed for English monolingual speakers. Therefore, some of the included words may be too difficult for second-language learners. For this reason, in this study lexical sophistication was calculated using the updated Target Word list.

Transcripts were run against the updated list using the CLAN program. The words that remained after "filtering" were deemed "sophisticated" for Dutch pupils in grade 1. For each transcript, CLAN calculated the number of different sophisticated words ($Types_{Sophisticated}$) and the number of total sophisticated words ($Tokens_{Sophisticated}$). In order to control for differences across

teachers in their amount of speech, the proportion of sophisticated tokens over the total number of tokens (excluding Dutch) produced was also calculated ($Sophisticated_{Proportion}$).

Syntactic complexity

In order to capture the syntactic complexity of teachers' language, each utterance was coded using a coding scheme based on the description of sentence complexity provided by Huttenlocher and his colleagues (2002) and the version of the coding scheme used by Justice, McGinty, Zucker, Cabell, and Piasta (2013). Each utterance was coded for sentence complexity using the following codes: (0) the utterance contained no clause (no verb structure), (1) the utterance contained one clause (simple, one verb structure), or (2) the utterance contained multiple clauses (complex, two or more verb structures). Table 3 provides an overview of the codes, including some examples of teacher utterances with corresponding codes. After coding each utterance, the proportion of complex utterances ($Complexity_{Proportion}$) was computed by dividing the number of complex utterances by the total number of utterances (including the 0 codes) to control for differences in amount of speech (Gámez and Lesaux 2012).

Table 3. Coding scheme for syntactic complexity with examples of teachers' utterances

Code	Teacher
(1) No Clause	On the table.
	Very well.
(2) Simple Clause	Why is that picture on the table?
(3) Complex Clause	We're gonna sing and we're gonna make noise like we always do even though there was a camera there.

Statistical analyses

To address the research questions, this study used Bayesian multilevel modeling. Since data were collected at both individual and teacher levels, it would have been inappropriate to use traditional methods of regression or ANOVA because of the violation of the independence assumption and the nested nature of the data. This is because when the assumption of independence is violated standard errors are underestimated which may lead to significant results when they are not actually there (Hox, Moerbeek, and van de Schoot 2010). Investigating the connection between individual- and contextual-level settings is known in the literature as hierarchical linear modeling (HLM) (Raudenbush and Bryk 2002) or multilevel modeling (MLM) (Goldstein 2003). In contrast to traditional ANOVA approaches, MLM estimation does not require balanced data and utilizes all available information in an unbalanced dataset such as the one used in this study. Todd, Crook, and Barilla (2005) argued that this particular characteristic makes it possible to detect relationships that might have previously gone undetected.

Furthermore, a Bayesian estimator was selected over the traditional maximum likelihood estimator to minimize the risk of producing untrustworthy estimates, particularly with small samples (e.g., Hox, van de Schoot, and Matthijsse 2012; McNeish and Stapleton 2016). MLMs with small sample sizes can be estimated with Bayesian Markov Chain Monte Carlo (MCMC) algorithm. MCMC does not utilize frequentist principles for estimation and, consequently, does not require large sample sizes for unbiased estimates (Gelman et al. 2013; Raudenbush and Bryk 2002). Although only a handful of studies have demonstrated this, Austin (2010), Browne and Draper (2006), and Stegmueller (2013) provided evidence that Bayesian estimates achieved unbiased estimates with lower numbers of clusters than likelihood methods even when fewer than ten clusters were present.

Ntzoufras (2009) argues that the Bayesian regression model involving the use of MCMC method consists of three steps: (a) the construction of prior probability distributions; (b) the

determination of a likelihood function; (c) sampling for the previously specified posterior probability distributions. Weakly informative default prior choices for variances have been proposed in the past for Bayesian MLMs. A weakly informative prior is a relatively weak statement of prior knowledge which, in the presence of small sample data, will help stabilize estimation and shrink the unstable and potentially biased maximum likelihood estimates toward the prior mean (Hamra, MacLehose, and Cole 2013). For example, Gelman (2006) considered half-t prior distributions for scale parameters in hierarchical models. The author proposed this weakly informative default prior to replace the sensitive Inverse-Gamma 'non-informative' prior in order to have a limiting posterior distribution for hierarchical models.

The statistical analyses were performed with Mplus 8 with the specification to use weakly informative priors (Muthén and Muthén 2017) based on previous literature (see below). A total number of 100000 iterations was used to get convergence for the small data set (Muthén 2017). The Bayes Factor (BF) that was associated with each hypothesis was used as an indication of the amount of evidence that there is in the data for each particular hypothesis. The BF summarizes the relative evidence in the data for a particular informative hypothesis as compared to another hypothesis. So, for instance, if a particular hypothesis receives a BF of 5, this indicates that the evidence in favor of this particular hypothesis is five times stronger than for the hypothesis that it is being compared to (Jeffrey 1948). Further, the posterior estimates of parameters and 95% credibility intervals were obtained and the deviance information criteria (DIC) were used to evaluate the goodness-of-fit of the models. The DIC is regarded as the Bayesian equivalent to the traditional Akaike information criterion (AIC). According to Spiegelhalter, Best, Carlin, and van der Linde (2002), smaller values indicate better fit. Syntax codes used in Mplus for the analysis are reported in the Appendix.

Results

As a precursor to the Bayesian MLM, descriptive statistics of children's vocabulary performance and teachers' language input were examined and preliminary analyses were carried out on the main teacher variables under study. All outcome and predictor variables of the sample were examined for missing values. The missing-data percentages are reported in Table 4.

Table 4. Percentages of missing data

Variable	Percentage missing	*N*
PPVT 1	17.0	200
PPVT 2	22.8	186
Mother tongue mother	34.0	159
Family SES	38.2	149

Descriptive statistics

On average, only less than 2% of teachers' talk was in Dutch (1.53% of Tokens and 1.55% of Utterances). As a group, teachers relied most frequently on no clause ($M = 110.67$, $SD = 30.47$) and simple utterances ($M = 178$, $SD = 40.57$) with minimal use of complex sentences ($M = 36.33$, $SD = 22.34$) (see Figure 1). The average amount of tokens and types teachers used was 1609 ($SD = 496.19$) and 365 ($SD = 88.13$), respectively.

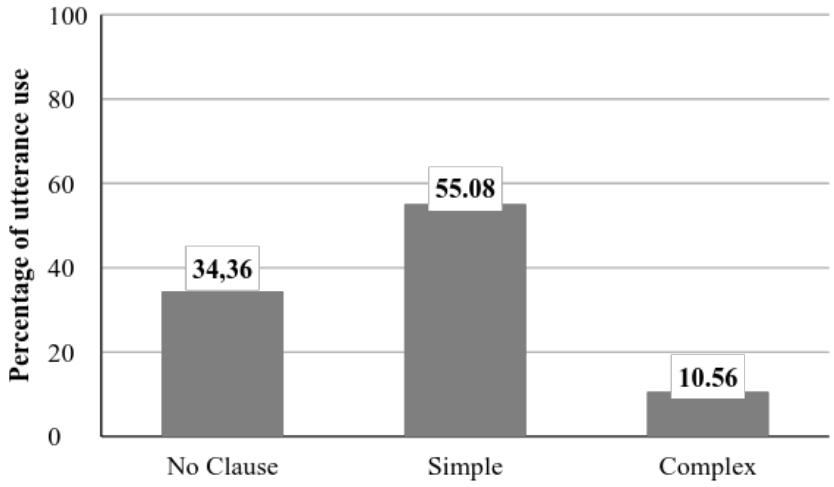

Figure 1. Teachers' syntactic complexity use.

Further, a Bayes test for correlations using the procedure of Wetzels and Wagenmakers (2012) revealed "substantial evidence" (Jeffreys 1948), favoring the hypothesis that there is a positive association between teachers' total amount of speech and their quality of speech ($Tokens_{Total}$ and $Types_{Sophisticated}$ $r = .78$, $BF_{10} = 11.69$, $N = 9$; $Utterances_{Total}$ and $Utterances_{Complex}$ $r = .76$, $BF_{10} = 9.17$, $N = 9$), underscoring the need to control for total amount of speech in the analyses. Evidence for a positive correlation was also found between $Utterances_{Total}$ and $Tokens_{Sophisticated}$ ($r = .68$, $BF_{10} = 4.46$). The correlations between all teacher variables are displayed in Table 5.

Table 5. Summary of the bivariate Bayesian correlations of all teacher variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Teaching experience	1						
(2) Teacher language proficiency	.48 [.59;.40]	1					
(3) $Complexity_{Proportion}$	.03 [.16; -.09]	.50 [.59;.39]	1				
(4) $Utterances_{Total}$	.46 [.55;.35]	.91 [.93;.89]	.45 [.55;.35]	1			
(5) $Tokens_{Total}$	.39 [.49;.28]	.80 [.84;.74]	.61 [.68;.52]	.93	1		
(6) $Sophisticated_{Proportion}$	.04 [.17; -.09]	.08 [.21; -.05]	-.13 [-.01; -.25]	.30 [.40;.17]	.38 [.48;.26]	1	
(7) Lexical diversity	.62 [.69;.53]	.58 [.66;.49]	.55 [.63;.46]	.73 [.78;.66]	.83 [.86;.78]	.48 [.38;.57]	1

Note. 95% credibility intervals are displayed in brackets.

Moreover, the preliminary analyses revealed a significant positive association between class mean scores of PPVT 1 and teachers' $Complexity_{Proportion}$ ($r = .58$, $BF_{10}>150$), $Sophisticated_{Proportion}$ ($r = .79$, $BF_{10}>150$) and Lexical Diversity ($r = .37$, $BF_{10}>150$), as well as with teachers' $Tokens_{Total}$ ($r = .60$, $BF_{10}>150$) and $Utterances_{Total}$ ($r = .45$, $BF_{10}>150$), indicating that children with higher vocabulary scores were enrolled in classrooms with teachers who used more words and utterances and more complex and diverse language.

Table 6 further shows substantial variation in children's raw scores on the PPVT measure. Comparing children's means of the PPVT scores at measurement 1 and 2 with a Bayesian paired t-test resulted in "decisive" evidence (BF = 163.70) in favor of the alternative hypothesis (there is a statistically reliable difference between the two means). This analysis thus indicated significant

improvement on receptive language skills from kindergarten to grade 1.

Table 6. Descriptive statistics of variables used in the analyses

Variable name	*N*	*M*	*SD*	Min	Max
Student-level statistics (Level 1)					
Age (in years)	241	6.76	.36	6	8
Family SES	149	2.81	.49	1	3
Mother's language					
Dutch	111				
English	8				
Fries	2				
Other	38				
Gender					
Female	113				
Male	128				
PPVT 1	200	56.88	22.61	12	108
PPVT 2	186	71.29	20.91	17	127
Teacher-level statistics (Level 2)					
Teaching experience (in years)	9	17.67	15.89	3	40
English proficiency (0 = Nonnative; 1= Near native; 2 = Native)	9	.89	.78	0	2
$Utterances_{Total}$	9	325.11	81.99	224	437
$Tokens_{Total}$	9	1454.78	492.04	720	2030
$Complexity_{Proportion}$	9	.11	.52	.03	.21
$Sophisticated_{Proportion}$	9	.04	.02	0.01	0.07
Lexical Diversity	9	7.77	.89	6.52	8.97

The relation between teachers' input and children's vocabulary

A fully unconditional model, with children's raw PPVT 2 scores as the outcome variable, showed significant variability within children (χ^2= 378.209; 95% CI [308.89-470.25]) and among children within teachers (χ^2= 164.66; 95% CI [41.85-821.01]). The credibility interval for the intra-class correlation (ICC = 0.30; 95% CI [0.096- 0.686]) was also bounded away from zero, indicating that a multilevel model is appropriate and needed. The next step was thus to build a sequence of two-level Bayesian MLMs including several covariates.

To answer research question (1), a level-1 model was built, with children's raw PPVT 1 scores added as group-mean centered and age as a grand-centered covariate, while gender, SES, and mother's language were uncentered, because the value zero was meaningless on their scale. The intercept was allowed to vary randomly by teacher, but all other effects were fixed. Adding the level 1 predictors reduced the Bayesian deviance of the model and the between-clusters variance (χ^2= 144.94; 95% CI [20.09-815.40], suggesting a better fit. As Table 7 shows, SES and scores on PPVT 1 significantly impacted children's vocabulary scores at measurement 2, whereas their native language and age were not significant predictors. Gender almost reached statistical significance so it was included in the next model.

Table 7. Level 1 model for children's vocabulary knowledge

	Posterior estimate	Posterior SD	*p*-MCMC	I-95% CI	U-95% CI
Intercept	41.86	14.52	.002	12.97	70.27
Gender	5.90	3.60	.052	-1.18	12.95
SES	8.54	4.33	.024	0.11	17.18
Age	6.98	6.58	.144	-5.98	19.82
Mother's language	1.10	1.60	.248	-2.04	4.24
PPVT 1	0.353	0.09	<.001	0.17	0.53
Deviance (DIC): -5271.743					

Note. *N* = 101. SD = standard deviation, CI = credibility interval.

To answer research questions (2) through (3), a level-2 model was built that included the teacher –level variables. To avoid convergence and collinearity issues, two separate models were used to examine the impact of teachers' input at the word level and utterance level. The word-level model included teachers' $Tokens_{Total}$ along with Lexical Diversity (modelled as a construct composed of $Sophisticated_{Proportion}$ and Guiraud's index to reduce collinearity). The utterance-level model included $Utterances_{Total}$ and $Complexity_{Proportion}$. The control variables Teaching Experience and English Proficiency were also included. All predictors except for English Proficiency were grand-mean centered. For a description of the final models see Appendix.

For Lexical Diversity, the selection of weakly informative priors was guided by the studies of Gamez and Lesaux (2012), Gamez (2015), and Bowers and Vasilyeva (2011), namely, a mean of 0.4 and a standard deviation of 0.05 was chosen. For $Complexity_{Proportion}$, the priors were chosen based on the studies of Gamez and Lesaux (2012) and Gamez et al. (2017), leading to a mean of 0.5 and a standard deviation of 0.05. The priors for $Tokens_{Total}$ were selected based on the same studies with M = 1.500 and SD = 1. The priors for $Utterances_{Total}$ could not be determined due to the lack of clear information provided in the studies.

Word-level model

As Table 8 shows, the results of the word-level model revealed that with all student-variables controlled for, teachers' lexical diversity significantly predicted children's vocabulary scores, whereas the total amount of tokens, teachers' English proficiency and teaching experience were not significant predictors. Adding the teacher input predictors accounted for 24% of the student-level variance in outcome. To assess whether the effects of lexical diversity were different depending on teachers' experience and English proficiency, an interaction of these variables were added to the model. Teachers' experience ($\beta = 1.12, p = .228$) and language proficiency ($\beta = 5.79, p = .328$) did not impact

the relation between lexical diversity and vocabulary scores.

Table 8. Results of word-level MLM on children's vocabulary knowledge

	Posterior estimate	Posterior SD	*p*-MCMC	I-95% CI	U-95% CI
Intercept	59.12	76.88	.152	-210.18	78.84
Gender	5.94	3.52	.045	-0.97	12.86
SES	6.66	3.99	.047	-1.16	14.48
PPVT 1	0.36	0.09	<.001	0.18	0.53
Lexical diversity	0.40	0.22	.037	-0.04	0.84
$Tokens_{Total}$	0.01	0.08	.203	-0.11	0.20
Teaching experience	1.14	4.87	.096	-0.33	15.48
English proficiency	1.47	0.99	.070	-0.48	3.43
Deviance (DIC): 883.203					

Note. $N = 101$. SD = standard deviation, CI = credibility interval.

Utterance-level model

As seen in Table 9, the results of the utterance-level model showed that with all other student-variables controlled for, teachers' use of complex utterances significantly impacted children's vocabulary scores, whereas total amount of utterances, teachers' English proficiency and teaching experience were not significant predictors. Adding the teacher input predictors accounted for 23% of the student-level variance in outcome. To assess whether the effects of lexical diversity were different depending on teachers' experience and English proficiency, an interaction of these variables were added to the model. The results showed that teachers' experience ($\beta = 0.80, p = .487$) and language proficiency ($\beta = 3.51, p = .172$) did not impact the relation between syntactic complexity and vocabulary scores.

Table 9. Results of utterance-level MLM on children's vocabulary knowledge

	Posterior estimate	Posterior SD	*p*-MCMC	I-95% CI	U-95% CI
Intercept	-12.29	26.82	.283	-57.51	45.70
Gender	5.89	3.49	.045	-0.91	12.81
SES	6.14	3.95	.059	-1.57	13.94
PPVT 1	0.39	0.09	<.001	0.21	0.56
$Complexity_{Proportion}$	0.50	0.22	.012	0.06	0.94
$Utterances_{Total}$	0.06	0.23	.363	-0.39	0.51
Teaching experience	0.01	1.25	.495	-2.66	2.31
English proficiency	1.51	1.00	.065	-0.46	3.47
Deviance (DIC): 894.147					

Note. $N = 101$. SD = standard deviation, CI = credibility interval.

As noted above, Bower and Vasilyeva (2011) found language level proficiency of children to play an important role in the relation between linguistic input and children's vocabulary outcomes. The positive correlation discussed above, suggesting that children with higher vocabulary scores were enrolled in classes with teachers who used more complex syntax and more lexically diverse language, led to the question whether teachers' use of more complex language may only benefit children with advanced English vocabulary knowledge. Thus, post hoc analyses were ran to further investigate whether teachers' syntactic complexity and lexical diversity would have a different influence on different subsets of the group, as a function of their vocabulary knowledge. Specifically, the data was split according to children's PPVT 2 scores (see Table 10).

Table 10. Children's distribution according to language proficiency

Language proficiency	Percentage	*N*
High & Average	71.8	173
Low	28.2	68

Two groups were created consisting of those children who were deemed as having either high or average (raw scores above 63) or low (raw scores below 62) proficiency levels. Although the number of Level 1 observations decreased substantially (see Table X), an utterance-level HLM with two levels was used because these two new samples were still evenly distributed across classrooms (M = 11.12, SD = 4.58 and M = 22.63, SD = 7.77). $\text{Complexity}_{\text{Proportion}}$ and Lexical Diversity were included as the main predictors, along with the student-level covariates SES, Gender, and PPVT 1. The results revealed a non-significant positive association (β= 52.68, SD = 78.12, df = 6, p = .204) between the proportion of complex utterances and receptive vocabulary scores for children with advanced or average language knowledge, and a non-significant negative association (β= -25.18, SD = 65.64, df = 6, p = .315) for children with a lower proficiency level.

Similarly, the post-hoc analysis resulted in a non-significant positive association (β= 1.61, SD = 11.74, df = 6, p = .475) between lexical diversity and vocabulary scores for children with high proficiency levels, whereas for children with poor vocabulary knowledge the coefficient was negative (β= -0.01, SD = 7.33, df = 6, p = .500). While the non-significant results are probably due to the reduction in sample size, it is important to note, though, that the obtained coefficients for the two subgroups move in opposite directions. This suggests that teachers' complex and diverse language use may have a negative effect on the vocabulary outcomes of children with lower vocabulary abilities.

Discussion

The aim of the present study was to investigate the quantity and quality of children's classroom-based language exposure in bilingual primary schools and the relation of teachers' linguistic input to children's receptive vocabulary scores. The main finding suggests that the quality of the English language used by teachers is a significant source of influence on children's vocabulary knowledge.

The first research question of this study concerned investigating the relation between children's individual characteristics and their vocabulary knowledge. The findings indicated that individual factors played an important role in children's receptive vocabulary knowledge. Specifically, gender, socioeconomic status and prior knowledge were significant predictors for their language outcomes. These results support previous research in this area suggesting that internal factors have a great influence on children's second language learning (Cook 2016; Paradis 2011).

The second research question addressed the relation between quantity of input and vocabulary outcomes. This study showed that the total amount of talk (i.e. amount of tokens and utterances) did not significantly influence children's vocabulary knowledge. This is inconsistent with other classroom and home-based studies that found significant effects of amount of talk such as Bowers and Vasilyeva (2011) and Huttenlocher et al. (2002) but in line with the recent literature on classroom-based language input (Gamez 2015; Gamez et al. 2017, Gamez and Lesaux 2012).

Concerning the relation between children's vocabulary knowledge and the quality of teacher input, measured in this study in terms of lexically diverse speech and complex syntax, the analyses showed that after controlling for children's initial vocabulary scores and their demographics, teachers' use of a diverse set of sophisticated vocabulary and complex syntax were positively related to children's word-learning. These results are in agreement with those obtained by Aukrust and Rydland

(2011), Dickinson and Porche (2011) Gamez and Lesaux (2012), Gamez (2015) and Gamez et al. (2017).

However, follow-up analyses further showed that teacher's syntactic complexity and lexically diverse language differentially benefited children's vocabulary outcomes. Specifically, teachers' complex syntax and use of diverse vocabulary had a negative influence on those children with lower English language proficiency (i.e. raw scores below 62), similarly to the findings of Bowers and Vasilyeva (2011) but contrary to Gamez and Lesaux (2012) and Gamez (2015). Importantly, the way syntactic complexity was operationalized differed across these studies. While Bowers and Vasilyeva (2011) and Gamez (2015) used the Mean Length of Utterance measure, Gamez and Lesaux (2012) coded teachers' syntax by means of a coding scheme, similarly to the one employed in the current study. Given this, differences in results cannot be attributed to different documentation methods. In fact, it could be that in the present study teacher's speech may have been too complex for low-English proficiency children to aid their word learning. It has been proposed that exposure to multiple exemplars of syntactic structures provides L2 children with a rich source of information (Onnis 2012), which in the present study seems to have benefited L2 children with advanced skills. Yet, L2 children with poor vocabulary knowledge may still struggle to successfully use syntactic cues in their second language, especially at early stages.

Important to note is that only the study of Bowers and Vasilyeva investigated the relation between *receptive* vocabulary knowledge and teachers' input; the other studies documented children's learning outcomes in terms of *expressive* language abilities. Moreover, whereas Bowers and Vasilyeva relied only on the number of word types to assess lexical diversity, the present study used a construct consisting of two separate but related measures (Guiraud's index and proportion of sophisticated tokens), thus also accounting for the frequency of words. The negative relations suggest that L2

children need first to acquire a more basic vocabulary before benefitting from the exposure to more sophisticated language.

Another important result was the failure to find any relation between teachers' language proficiency and their use of diverse and complex language, meaning that being a native speaker and having more teaching experience does not necessarily lead to different language use, which is in contrast to the results obtained by Unsworth et al. (2014). However, it is important to note that the teachers in the present study had rather high levels of English proficiency, contrary to Unsworth et al (2014). This finding suggests that teachers may have more of a personal style of speaking, which could be related to the distribution of children across classrooms. Namely, the analyses indicated that children with higher vocabulary knowledge were enrolled in classrooms with teachers who used more diverse and complex language, which raises the possibility that some teachers may have fine-tuned their speech to meet the needs of their children, while others failed to do so, leading to the negative associations described above.

Limitations and future research

These findings should be interpreted in light of several limitations. Firstly, due to some logistical issues, this study used data from a small sample size making generalization difficult and relied on video-observations of only one lesson. Also, given the nature of the study, the results cannot be used to suggest any causal relationships. What they provide is interesting evidence of variations in the quality of linguistic input in primary bilingual schools. In order to draw firmer conclusions about the impact of the quality and quality of English exposure on L2 children's vocabulary learning, more studies are needed that conduct similarly linguistic analyses with larger samples of classrooms.

Secondly, this study focused on the classroom average of linguistic input and did not examine individual children's speech or teachers' talk to individual children. Future research could investigate bidirectional relations between teachers' and children's speech, for instance, by analyzing teacher-

child turn taking during small group activities. It is possible that teachers modify their speech in response to individual children, using either more simplified or complex language depending on children's proficiency.

Not including children's native language knowledge as a control variable could also be regarded as a potential limitation. By the time the analyses were conducted, these data were still being processed and could therefore not be considered in this study. First language vocabulary skills have been found to impact second language vocabulary knowledge in various ways (for review see Proctor et al. 2006). Moreover, this study only relied on one standardized method for receptive vocabulary assessment, the Peabody Picture Vocabulary Test, which is actually designed for monolinguals and could potentially contain culturally biased items (Sweet, Csillag, and Lebron 2007). Further work involving a more comprehensive assessment of L2 children's vocabulary skills in both languages should allow for a more systematic investigation.

Furthermore, a missing link in studies on classroom-based language exposure is a focus on the peer contributions to the language environment. Higher peer expressive language abilities have been found to be positively associated with children's development of receptive and expressive language during pre-kindergarten (Mashburn et al. 2009). Future studies addressing the influence of classroom-based language exposure should also take into account children's language abilities within a classroom as a potential source of variation and its influence on teachers' linguistic behavior.

Although this study has certain methodological limitations and thus the findings should be viewed as suggestive rather than definitive, it outlines a potentially productive direction for future research. Overall, the finding that teachers' personal speaking styles may account for variation in linguistic input across classrooms has significant implications for the design of effective learning environments for all children, but particularly for those with underdeveloped language abilities. It

could mean, for instance, that professional development efforts could focus on developing teachers' complex and diverse language use in deliberate and meaningful ways in the service of promoting the language development of all children. This could be done by offering extra support to children with lower proficiency levels, for example by organizing group activities in which teachers use more simplified language to first ensure basic vocabulary learning. Also, encouraging verbal interactions with more skilled peers may result in qualitative language exposure.

Conclusion

The examination of both internal and external factors associated with variability in language learning may have critical implications for the improvement of language outcomes of children enrolled in bilingual education programs. The present study is a first step in examining the direct relation between teacher language input and early word learning for Dutch children from bilingual classrooms while accounting for their demographics. The findings suggest that exposure to high-quality language has beneficial effects for children's vocabulary outcomes. However, this study also found a differential influence of teachers' language use as a function of children's English language proficiency. Future work should extend this line of investigation by carrying out empirical research involving larger samples and exploring multiple sources of language input. This line of research may lead to the design of more effective primary school curriculum and teacher training programs.

References

Admiraal, Wilfried, Gerard Westhoff, and Kees de Bot. 2006. "Evaluation of Bilingual Secondary Education in the Netherlands: Students' Language Proficiency in English." *Educational Research and Evaluation* 12 (1): 75–93. doi:10.1080/13803610500392160.

Aukrust, Vibeke Grøver. 2007. "Young Children Acquiring Second Language Vocabulary in Preschool Group-Time: Does Amount, Diversity, and Discourse Complexity of Teacher Talk Matter?" *Journal of Research in Childhood Education* 22 (1): 17–37. doi:10.1080/02568540709594610.

Aukrust, Vibeke Grøver, and Veslemøy Rydland. 2011. "Preschool Classroom Conversations as Long-Term Resources for Second Language and Literacy Acquisition." *Journal of Applied Developmental Psychology* 32 (4): 198–207. doi:10.1016/j.appdev.2011.01.002.

Austin, Peter C. 2010. "Estimating Multilevel Logistic Regression Models When the Number of Clusters Is Low: A Comparison of Different Statistical Software Procedures." *The International Journal of Biostatistics* 6 (1). doi:10.2202/1557-4679.1195.

Birdsong, David. 1999. *Second Language Acquisition and the Critical Period Hypothesis*. London: Lawrence Erlbaum Associates.

Bowers, Edmond P, and Marina Vasilyeva. 2011. "The Relation between Teacher Input and Lexical Growth of Preschoolers." *Applied Psycholinguistics* 32 (1): 221–41. doi:10.1017/S0142716410000354.

Brown, Roger. 1973. *A First Language: The Early Stages*. London: George Allen & Unwin Ltd.

Browne, William J., and David Draper. 2006. "A Comparison of Bayesian and Likelihood-Based Methods for Fitting Multilevel Models." *Bayesian Analysis* 1 (3). International Society for Bayesian Analysis: 473–514. doi:10.1214/06-BA117.

Cantone, Katja Francesca. 2007. *Code-Switching in Bilingual Children*. Springer.

Celaya, Maria Luz. 2012. "'I Wish I Were Three!': Learning EFL at an Early Age." In *New Trends in Early Foreign Language Learning: The Age Factor, CLIL and Languages in Contact. Bridging Research and Good Practices*, edited by M. González Davies and A. Taronna, 2–12. Cambridge Scholars' Publishing.

Cenoz, Jasone. 2013. "Discussion: Towards an Educational Perspective in CLIL Language Policy and Pedagogical Practice." *International Journal of Bilingual Education and Bilingualism* 16 (3). Taylor & Francis Group: 389–94. doi:10.1080/13670050.2013.777392.

Chall, Jeanne S, and Edgar Dale. 1995. *Readability Revisited: The New Dale-Chall Readability Formula*. Brookline Books.

Chaudron, Craig. 1988. *Second Language Classrooms: Research on Teaching and Learning*. Cambridge: Cambridge University Press.

Cook, Vivian. 2016. *Second Language Learning and Language Teaching / Vivian Cook. - Version Details - Trove*. 5thed. London: Routledge.

Council of Europe. 2001. "The Common European Framework of Reference for Languages." Cambridge University Press. https://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf.

Coyle, Do, Philip Hood, and David Marsh. 2010. *CLIL: Content and Language Integrated Learning*. Cambridge University Press.

Dalton-Puffer, Christiane, Tarja Nikula, and Ute Smit. 2010. *Language Use and Language Learning in CLIL Classrooms*. Edited by Christiane Dalton-Puffer, Tarja Nikula, and Ute Smit. Vol. 7. AILA Applied Linguistics Series. Amsterdam: John Benjamins Publishing Company. doi:10.1075/aals.7.

DeKeyser, Robert M. 2000. "The Robustness of Critical Period Effects in Second Language Acquisition." *Ssla* 22: 499–533. doi:0272-2631/00.

Dickinson, David K., and Michelle V. Porche. 2011. "Relation Between Language Experiences in Preschool Classrooms and Children's Kindergarten and Fourth-Grade Language and Reading Abilities." *Child Development* 82 (3): 870–86. doi:10.1111/j.1467-8624.2011.01576.x.

Driessen, Gert, Evelien Krikhaar, Rick De Graaff, Sharon Unsworth, Bianca Leest, Karien Coppens, and Janice Wierenga. 2016. "Evaluation Pilot Bilingual Primary Education." *ITS, Radboud Universiteit Nijmegen*. https://www.nuffic.nl/en/files/documents/evaluation-pilot-bilingual-primary-education.pdf.

Dunn, Lloyd, and Douglas Dunn. 2007. *Peabody Picture Vocabulary Test*. 4thed. Pearson Assessments.

Ellis, Rod. 2004. "Individual Differences in Second Language Learning." In *The Handbook of Applied Linguistics*, 525–51. Oxford, UK: Blackwell Publishing Ltd. doi:10.1002/9780470757000.ch21.

EP-Nuffic. 2017. *Streefwoordenlijst Engels Voor Groep 1 En 2 Uit Het Tweetalig Primair Onderwijs (TPO)*. Accessed August 23. https://www.nuffic.nl/bestanden/documenten/streefwoordenlijst-engels-voor-groep-1-en-2-uit.pdf.

EP Nuffic. 2016. "Pilot Tweetalig Primair Onderwijs in Nederland." https://www.nuffic.nl/bestanden/documenten/pilottweetaligprimaironderwijsinnederland.pdf.

Francis, David, Nonie Lesaux, and Diane August. 2006. "Language of Instruction." *Developing Reading and Writing in Second-Language Learners: Lessons from the Report of the National Literacy Panel on Language-Minority Children and Youth*.

Gámez, Perla B. 2015. "Classroom-Based English Exposure and English Language Learners' Expressive Language Skills." *Early Childhood Research Quarterly* 31: 135–46.

doi:10.1016/j.ecresq.2015.01.007.

Gámez, Perla B., and Nonie K. Lesaux. 2012. "The Relation Between Exposure to Sophisticated and Complex Language and Early-Adolescent English-Only and Language Minority Learners' Vocabulary." *Child Development* 83 (4): 1316–31. doi:10.1111/j.1467-8624.2012.01776.x.

Gámez, Perla B., and Susan C. Levine. 2013. "Oral Language Skills of Spanish-Speaking English Language Learners: The Impact of High-Quality Native Language Exposure." *Applied Psycholinguistics* 34: 1–24. doi:10.1017/S0142716411000919.

Gámez, Perla B., Sabina Rak Neugebauer, Michael D. Coyne, D. Betsy McCoach, and Sharon Ware. 2017. "Linguistic and Social Cues for Vocabulary Learning in Dual Language Learners and Their English-Only Peers." *Early Childhood Research Quarterly* 40. Elsevier Inc.: 25–37. doi:10.1016/j.ecresq.2017.01.003.

Gelman, Andrew. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)." *Bayesian Analysis* 1 (3): 515–34. doi:10.1214/06-BA117A.

Gelman, Andrew, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. 2013. *Bayesian Data Analysis*. 3rded.

Genesee, Fred. 1983. "An Invited Article: Bilingual Education of Majority-Language Children: The Immersion Experiments in Review." *Applied Psycholinguistics* 4 (1): 1–46. doi:10.1017/S0142716400001739.

———. 1985. "Second Language Learning through Immersion: A Review of U.S. Programs." *Review of Educational Research* 55 (4): 541. doi:10.2307/1170246.

———. 2004. "What Do We Know about Bilingual Education for Majority Language Students." *Handbook of Bilingualism*, 547–76. doi:10.1002/9780470756997.ch21.

Goldenberg, Claude. 2008. “Teaching English Language Learners; What the Research Does-and Does Not-Say.pdf.” *American Educator*.

Goldstein, Harvey. 2003. *Multilevel Statistical Models*. 3rded. London: Edward Arnold.

Greene, Jay. 1997. “A Meta-Analysis of the Rosell & Baker Review of Bilingual Education Research.” *Bilingual Research Journal* 21 (2): 103–22. doi:10.1080/15235882.1997.10668656.

Guiraud, Pierre. 1960. *Problèmes et Méthodes de La Statistique Linguistique*. Paris: Presses Universitaires de France.

Hamra, Ghassan B., Richard F. MacLehose, and Stephen R. Cole. 2013. “Sensitivity Analyses for Sparse-Data Problems—Using Weakly Informative Bayesian Priors.” *Epidemiology* 24 (2): 233–39. doi:10.1097/EDE.0b013e318280db1d.

Hoff, Erika. 2006a. “Environmental Supports for Language Acquisition.” In *Handbook of Early Literacy Research*, edited by David Dickinson and Susan Neuman, 2nded., 163–72. New York: Guilford.

———. 2006b. “How Social Contexts Support and Shape Language Development.” *Developmental Review* 26 (1): 55–88. doi:10.1016/j.dr.2005.11.002.

Hoff, Erika, and Letitia Naigles. 2002. “How Children Use Input to Acquire a Lexicon.” *Child Development* 73 (2): 418–33. doi:10.1111/1467-8624.00415.

Hout, Roeland Van, and Anne Vermeer. 1988. “Spontane Taaldata En Het Meten Van Lexicale Rijkdom In Tweedetaalverwerving.” *Toegepaste Taalwetenschap in Artikelen* 32: 108–22. doi:10.1075/ttwia.32.07hou.

Houwer, Annick De. 2007. “Parental Language Input Patterns and Children’s Bilingual Use.” *Applied Psycholinguistics* 28 (3): 411–24. doi:10.1017/S0142716407070221.

Hox, Joop J., Rens van de Schoot, and Suzette Matthijsse. 2012. "How Few Countries Will Do? Comparative Survey Analysis from a Bayesian Perspective." *Survey Research Methods* 6 (2): 87–93. doi:10.18148/srm/2012.v6i2.5033.

Huttenlocher, Janellen. 1998. "Language Input and Language Growth." *Preventive Medicine* 27 (2): 195–99. doi:10.1006/pmed.1998.0301.

Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons. 1991. "Early Vocabulary Growth: Relation to Language Input and Gender." *Developmental Psychology* 27 (2). American Psychological Association: 236–48.

Huttenlocher, Janellen, Marina Vasilyeva, Elina Cymerman, and Susan Levine. 2002. "Language Input and Child Syntax." *Cognitive Psychology* 45 (3): 337–74. doi:10.1016/S0010-0285(02)00500-5.

Hüttner, Julia, and Angelika Rieder-Bünemann. 2007. "The Effects of CLIL Instruction on Children's Narrative Competence." *Vienna English Working Papers* 15(3). 41-46.

Jeffreys, Harold. 1948. *Theory of Probability*. Oxford: Clarendon Press.

Kanokpermpoon, Monthon. 2013. "Managing Working Memory in Language Instructions: An Overview of Cognitive Load Theory." *Thammasat Review* 16 (2): 93–108.

Lambert, Wallace E., and G. Richard Tucker. 1972. *Bilingual Education of Children; the St. Lambert Experiment*. Newbury House Publishers.

Leij, Aryan van der, Judith Bekebrede, and Mieke Kotterink. 2010. "Acquiring Reading and Vocabulary in Dutch and English: The Effect of Concurrent Instruction." *Reading and Writing* 23 (3): 415–34. doi:10.1007/s11145-009-9207-5.

Lervåg, Arne, and Vibeke Grøver Aukrust. 2010. "Vocabulary Knowledge Is a Critical Determinant of the Difference in Reading Comprehension Growth between First and Second Language

Learners." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 51 (5): 612–20. doi:10.1111/j.1469-7610.2009.02185.x.

Littlewood, William. 1984. *Foreign and Second Language Learning: Language-Acquisition Research and Its Implications for the Classroom*. Cambridge University Press.

Llinares, Ana, Tom Morton, and Rachel. Whittaker. 2012. *The Roles of Language in CLIL*. Cambridge University Press.

Macwhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing*. 3rded. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, Brian. 1995. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum.

Malvern, David, and Brian Richards. 2002. "Investigating Accommodation in Language Proficiency Interviews Using a New Measure of Lexical Diversity." *Language Testing* 19 (1): 85–104. doi:10.1191/0265532202lt221oa.

Marsh, David. 2002. *CLIL/EMILE - the European Dimension: Actions, Trends and Foresight Potential*. University of Jyväskylä.

Mashburn, Andrew J, Laura M Justice, Jason T Downer, and Robert C Pianta. 2009. "Peer Effects on Children's Language Achievement during Pre-Kindergarten." *Child Dev* 80 (3): 686–702. doi:CDEV1291 [pii]\r10.1111/j.1467-8624.2009.01291.x.

McNeish, Daniel M., and Laura M. Stapleton. 2016. "The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration." *Educational Psychology Review* 28 (2): 295–314. doi:10.1007/s10648-014-9287-x.

Muthén, Bengt. 2017. *Bayesian Analysis In Mplus: A Brief Introduction*. 3rded.

Muthén, Linda, and Bengt Muthén. 2017. *Mplus User's Guide*. 8thed. Los Angeles, CA: Muthén & Muthén.

Ntzoufras, Ioannis. 2009. *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: John Wiley & Sons.

Núñez, Begona, and Emma Dafouz. 2007. "Lecturing through the Foreign Language in a CLIL University Context: Linguistic and Pragmatic Implications." 3. Vienna English Working Papers.

Paradis, Johanne. 2011. "Individual Differences in Child English Second Language Acquisition: Comparing Child-Internal and Child-External Factors." *Linguistic Approaches to Bilingualism* 1 (3): 213–37. doi:10.1075/lab.1.3.01par.

Paradis, Johanne, Elena Nicoladis, Martha Crago, and Fred Genesee. 2011. "Bilingual Children's Acquisition of the Past Tense: A Usage-Based Approach." *Journal of Child Language* 38 (3): 554–578. doi:10.1017/S0305000910000218.

Pinker, Steven. 1994. *The Language Instinct*. New York: Harper Perennial Modern Classiscs.

Proctor, Patrick, Diane August, Maria S. Carlo, and Catherine Snow. 2006. "The Intriguing Role of Spanish Language Vocabulary Knowledge in Predicting English Reading Comprehension." *Journal of Educational Psychology* 98 (1): 159–69. doi:10.1037/0022-0663.98.1.159.

Quiroz, Blanca G., Catherine E. Snow, and Jing Zhao. 2010. "Vocabulary Skills of Spanish--English Bilinguals: Impact of Mother--Child Language Interactions and Home Language and Literacy Support." *International Journal of Bilingualism* 14 (4): 379–99. doi:10.1177/1367006910370919.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models : Applications and Data Analysis Methods*. 2nd ed. Sage Publications.

Rietveld, Toni, and Roeland Van Hout. 1993. *Statistical Techniques for the Study of Language and Language Behaviour*. Mouton de Gruyter.

Rolstad, Kellie, Kate Mahoney, and Gene Glass. 2005. "The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners." *Educational Policy* 19 (4): 572–94. doi:10.1177/0895904805278067.

Rossell, Christine H., and Keith Baker. 1996. "The Educational Effectiveness of Bilingual Education." *Research in the Teaching of English* 30 (1): 7–74. doi:10.2307/40171543.

Rowe, Meredith L. 2012. "A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development." *Child Development* 83 (5). Blackwell Publishing Ltd: 1762–74. doi:10.1111/j.1467-8624.2012.01805.x.

Rydland, Veslemøy, Vibeke Grøver Aukrust, and Helene Fulland. 2012. "How Word Decoding, Vocabulary and Prior Topic Knowledge Predict Reading Comprehension. A Study of Language-Minority Students in Norwegian Fifth Grade Classrooms." *Reading and Writing* 25 (2): 465–82. doi:10.1007/s11145-010-9279-2.

Scheele, Anna F, Paul Leseman, and Aziza Mayo. 2009. "The Home Language Environment of Monolingual and Bilingual Children and Their Language Proficiency." *Applied Psycholinguistics* 31 (1): 117-140. doi:10.1017/S0142716409990191.

Scovel, Thomas. 2000. "A CRITICAL REVIEW OF THE CRITICAL PERIOD RESEARCH." Annual *Review of Applied Linguistics* 20: 213–23. doi:10.1017/S0267190500200135.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4): 583–639. doi:10.1111/1467-9868.00353.

Stegmueller, Daniel. 2013. "How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches." *American Journal of Political Science* 57 (3): 748–61. doi:10.1111/ajps.12001.

Sweet, Emily, Julia Csillag, and Christopher Lebron. 2007. "Picture Perfect ? Validity and Cultural Bias of the PPVT-4," 30–31.

Todd, Samuel Y., T. Russell Crook, and Anthony G. Barilla. 2005. "Hierarchical Linear Modeling of Multilevel Data." *Journal of Sport Management* 19 (4): 387–403. doi:10.1123/jsm.19.4.387.

Torruella, Joan, and Ramon Capsada. 2013. "Lexical Statistics and Tipological Structures: A Measure of Lexical Richness." *Procedia - Social and Behavioral Sciences* 95. Elsevier B.V.: 447–54. doi:10.1016/j.sbspro.2013.10.668.

Unsworth, Sharon. 2016. "Quantity and Quality of Language Input in Bilingual Language Development." *Bilingualism across the Lifespan: Factors Moderating Language Proficiency*, 103–22. doi:10.1037/14939-007.

Unsworth, Sharon, Liv Persson, Tineke Prins, and Kees De Bot. 2014. "An Investigation of Factors Affecting Early Foreign Language Learning in the Netherlands." *Applied Linguistics* 45 (5). Applied Linguistics. doi:10.1093/applin/amt052.

Vermeer, Anne. 2001. "Breadth and Depth of Vocabulary in Relation to L1/L2 Acquisition and Frequency of Input." *Applied Psycholinguistics* 22 (2): 217–34. https://www.cambridge.org/core/journals/applied-psycholinguistics/article/breadth-and-depth-of-vocabulary-in-relation-to-l1l2-acquisition-and-frequency-of-input/6A7CD827022283582903ED690EF4EE06.

Verspoor, Marjolijn, Kees de Bot, and Xiaoyan Xu. 2015. "The Effects of English Bilingual Education in the Netherlands." *Journal of Immersion and Content-Based Language Education* 3 (1): 4–27. doi:10.1075/jicb.3.1.01ver.

Weisleder, Adriana, and Anne Fernald. 2013. "Talking to Children Matters - Early Language Experience Strengthens Processing and Builds Vocabulary." *Psychological Science* 24 (11). SAGE Publications: 2143–52. doi:10.1177/0956797613488145.

Weizman, Zehava, and Catherine Snow. 2001. "Lexical Input as Related to Children's Vocabulary Acquisition: Effects of Sophisticated Exposure and Support for Meaning." *Developmental Psychology*. doi:10.1037/0012-1649.37.2.265.

Wetzels, Ruud, and Eric-Jan Wagenmakers. 2012. "A Default Bayesian Hypothesis Test for Correlations and Partial Correlations." doi:10.3758/s13423-012-0295-x.

Willig, Ann. C. 1985. "A Meta-Analysis of Selected Studies on the Effectiveness of Bilingual Education." *Review of Educational Research* 55 (3): 269–317. doi:10.3102/00346543055003269.

———. 1987. "Examining Bilingual Education Research Through Meta-Analysis and Narrative Review: A Response to Baker." *Review of Educational Research* 57 (3): 363–76. doi:10.3102/00346543057003363.